M.S. THESIS

VISUAL ATTENTION PROBABILITY MODEL FOR STEREOSCOPIC VIDEOS ESTIMATED USING STATISTICAL DESIGN OF EXPERIMENTS

통계적 실험 계획법을 이용하여 추정한 삼차원 동영상의 시각 주의 확률 모델

BY

KIM BO-EUN

FEBRUARY 2015

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

M.S. THESIS

VISUAL ATTENTION PROBABILITY MODEL FOR STEREOSCOPIC VIDEOS ESTIMATED USING STATISTICAL DESIGN OF EXPERIMENTS

통계적 실험 계획법을 이용하여 추정한 삼차원 동영상의 시각 주의 확률 모델

BY

KIM BO-EUN

FEBRUARY 2015

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

VISUAL ATTENTION PROBABILITY MODEL FOR STEREOSCOPIC VIDEOS ESTIMATED USING STATISTICAL DESIGN OF EXPERIMENTS

통계적 실험 계획법을 이용하여 추정한 삼차원 동영상의 시각 주의 확률 모델

지도교수 김 태 정

이 논문을 공학석사 학위논문으로 제출함

2014년 11월

서울대학교 대학원

전기컴퓨터 공학부

김 보 은

김보은의 공학석사 학위 논문을 인준함

2014년 11월

위 원 장: _____

부위원장: _____

위 원: _____

Abstract

Viewers of videos are likely to absorb more information from the part of the screen that attracts visual attention. This fact has lead to the visual attention models that is being used in producing and evaluating videos. In this paper, we investigate the factors that are significant to visual attention and the mathematical form of the visual attention model, and then estimate the visual attention probability using the statistical design of experiments. The analysis of variance (ANOVA) verifies that the motion velocity, distance from the screen, and amount of defocus blur are the factors that strongly affect human visual attention. Using the response surface modeling (RSM), we create a visual attention score model that concerns the three factors, and from which model we calculate the visual attention probabilities (VAPs) of image pixels. The VAPs are directly applied to existing gradient based 3D effect perception measurement. By giving weights according to our VAPs, more accurate measurement is possible. The performance of the proposed measurement is assessed by comparing them with subjective evaluation as well as with existing methods. The comparison verifies that the proposed measurement outperforms the existing ones.

keywords: Visual attention probability, stereoscopic video, Statistical design of experiments, 3D effect perception measurement, Defocus blur estimation
student number: 2013-20759

Contents

Abstract i

Contents ii

List of Figures iv

List of Tables v

1 Introduction 1

2 Visual Attention Probability (VAP) 3

2.1 Previous Studies and Motivation 3

2.2 Factors that Influence the Visual Attention 5

2.3 Statistical Design of the Experiment and the Visual Attention Score Model 5

2.3.1 Experiment Design 5

2.3.2 Analysis of Variance (ANOVA) 7

2.3.3 Response Surface Modeling (RSM) 9

2.4 Visual Attention Probability and Its Application 15

2.4.1 Visual Attention Probability 15

2.4.2 Application to Movie Frames 17

3 3D Effect Perception Measurement Using the VAP **27**

3.1 Previous Studies 27

3.2 Gradient Method with the VAP 28

3.3 Experiment Results 30

4 Conclusion **33**

Abstract in Korean **37**

List of Figures

2.1 Data points. 8
2.2 Visual Attention Score (VAS) graphs - velocity (V) fixed. 12
2.3 Visual Attention Score graphs - Distance From the screen (DFS) fixed. 13
2.4 Visual Attention Score graphs - Defocus Blur Amount (DBA) fixed. . 14
2.5 In-focus distance and DOF. 20
2.6 Example graphs of DBA according to distance from the screen in the specific in-focus distances. 21
2.7 Result images for test sequence 1. 22
2.8 Result images for test sequence 2. 22
2.9 Result images for test sequence 3. 23
2.10 Result images for test sequence 4. 23
2.11 Result images for test sequence 5. 24
2.12 Result images for test sequence 6. 24
2.13 Result images for test sequence 7. 25
2.14 Result images for test sequence 8. 25
2.15 Result images for test sequence 9. 26
2.16 Result images for test sequence 10. 26

3.1 Visual Sensitivity Kernel (VSK). 30
3.2 Scatter plot of measurements and subjective evaluation results. 32

List of Tables

2.1 Experimental environments 6
2.2 Factors and levels 7
2.3 ANOVA table 8
2.4 Mean of visual attention scores at each data point 10

Chapter 1

Introduction

In recent years, with the growth of the display industry, new technology-based products were released such as 3D TVs, high-definition and large-screen TVs, and curved TVs. Accordingly, the video contents were varied, and researches on video-making and evaluation methods became important. In the video-making or evaluation stage, it would be of great help to estimate the part of the image the viewer visually pays attention to among the other parts on the screen. When people stare at the screen, they cannot see the entire screen with the same visual sensitivity. Their visual sensitivity is keener at the points closer to the visually fixed-at point [1]. Therefore, viewers are likely to accept information more significantly from the part visually concentrated on. Using this fact, the visual attention model is being used in studies such as on video quality assessment [2].

In this paper, we estimate the visual attention probability (VAP) model in videos using the statistical design of experiments to investigate which factors are significant to visual attention and what would be the concrete form of the visual attention model.

In the second part of this paper, the VAP model is applied to the three-dimensional (3D) effect perception measurement method for stereoscopic videos. Since the 3D

movie Avatar was released in 2009, 3D movies have become a trend in the film industry. Also, 3D-watching devices have become common, as 3D functions are included in most of the latest TV models. It is very important to control the 3D effect when producing 3D videos. Film makers make depth charts before shooting, considering the story flow and the impact of the scenes. Likewise, when evaluating 3D videos, it is important to evaluate the degrees of the 3D effect perceived by the viewers. If we could measure the 3D effect from the video information, we could cut the evaluation time and cost [13]. Conversely, we can also use the measurement to adjust the 3D effect of the scenes when making videos.

In this paper, we aim to measure the 3D effects perceived by viewers of different scenes of stereoscopic videos. We propose the VAP model application method for the conventional 3D effect measurement method using the depth image to improve the performance. To measure the 3D effect more accurately, high weights were given to the parts with a high VAP.

Chapter 2

Visual Attention Probability (VAP)

2.1 Previous Studies and Motivation

Visual attention models have been studied previously. Park first combined skin color information with Itti's model, which used color, intensity, and the orientation of the intensity to make a 2D visual attention model, and he combined a disparity factor to propose a visual attention model of a 3D still image [4-6]. Park multiplied a nonlinear function by a 2D model with a heavy weight in the comfortable 3D viewing disparity range.

Kim proposed the visual attention model for 3D videos by adding a depth factor to Itti's model [2][4]. He made functions for a depth factor and a movement factor, and added them up after multiplying them by different weights. The function of the movement factor (T) is the linearly normalized function of S·m.

$$T = \psi(S \cdot m) \tag{2.1}$$

S is the result of Itti's 2D model, and m is the motion vector magnitude. ψ is the

linear normalization function that adjusts the value of the S·m from 0 to 255. For the function of the depth factor (D), original depth value (d) is linearly normalized, and the normalized values are transformed into zero, except for the top 15%.

$$D = P(\psi(d)) \tag{2.2}$$

$$P(a) = \begin{cases} a & , if\ a > D_{th} \\ 0 & otherwise \end{cases} \tag{2.3}$$

$$D_{th} = \big(max(D) - min(D)\big) \times 0.15 \tag{2.4}$$

Kim made the 3D Visual Attention (3DVA) model by multiplying different weights by the 2D model S, movement factor T, and depth factor D, and by linearly combining the three weighted terms. He set the values of w_s, w_t,w_d, and empirically at 0.2, 0.32, and 0.48, respectively.

$$3DVA = w_s S + w_t T + w_d D \tag{2.5}$$

In this paper, we estimate the visual attention probability (VAP) model of videos through experiments. We analyzed which factors are actually significant to visual attention and how the visual attention degree changes according to the significant factors. We scored the visual attention at discrete data points through subject evaluations. By modeling with these scores, we estimated the visual attention scores (VASs) in a continuous range. With this score model, we determined the probability that a viewer will concentrate on each pixel in a video frame (the VAP). Chapter 2.2 describes the factors that influence the visual attention. Chapter 2.3 describes the design of the experiments to obtain the VAS model and its results. Finally, Chapter 2.4 describes the VAP.

2.2 Factors that Influence the Visual Attention

For a 2D still image, color and intensity are considered the factors that significantly influence the visual attention. Itti used color, intensity, and the orientation of the intensity characteristics [4]. In the case of 3D videos, depth and movement are the factors that influence the visual attention. Kim composed a model that includes these two factors [2]. In this paper, we use the term 'distance from the screen' (DFS) instead of 'depth'.

In this paper, we estimate the VAP model of videos. The estimated model is suitable for both 2D and 3D videos. We used two factors that were proposed by Kim and the defocus blur amount (DBA) information. Velocity (V) is expected to be a significant factor because the targets are videos. The DFS information is used to respond to 3D videos. In addition, the focus is expected to significantly influence the visual attention. When we take a picture, we focus the camera on the main objects at a certain distance, and they come out clearly on the photograph. The visual attention degree would be higher in the region in focus than in the region that is out of focus.The DBA reflects whether the region is in focus or not. Consequently, these three factors–the V, DFS, and DBA–are expected to significantly influence the visual attention, so the experiment was conducted with such factors.

2.3 Statistical Design of the Experiment and the Visual Attention Score Model

2.3.1 Experiment Design

The visual attention scores (VASs) were obtained through an experiment at different levels of each of these three factors: the velocity (V), distance from the screen (DFS), and defocus blur amount (DBA). The levels of factor were selected from the range of

numbers usually appear in real 3D movies. The levels are summarized in Table 2.2. They are expressed as -α, -1, 0, +1, and + α for convenience, but it does not mean that the difference between the level values is proportional to the difference between the level expressions. For generalization, the unit of the V was converted from the pixel/frame length on the image to m/s on the screen. The experiment environment in Table 2.1 was used as a standard for the conversion. The DFS was expressed as a negative value when the object was behind the screen, a positive value when the object stuck out in front of the screen, and zero when the object was on the screen. Gaussian kernels are generally used in order to express defocus blurs. In the case of the DBA factor, Gaussian filters with different standard deviations were used to produce the experiment videos. Such standard deviation values are arranged at the DBA level in Table 2.2. The unit of the DBA level used for the modeling was m, which was converted to such from the pixel.

Table 2.1: Experimental environments

Test video resolution	1920×1080 stereoscopic videos
Frame rate	24 fps
Display	16:9 / 46 inch 3DTV
Viewing distance	3m

We cannot obtain the interactions between the factors if we construct three experiments related to each factor, respectively, to examine the effect of the three factors of visual attention [8]. Therefore, in this paper, we constructed an experiment related to all three factors at the same time. The points in Figure 2.1 are the data points from the combinations of the levels of the factors. The vertical, horizontal, and up-down axes are the level variation axes of the three factors, respectively. Eight black points were required for the Analysis of Variance (ANOVA), and seven gray points were additionally used for the Response Surface Modeling (RSM).

Table 2.2: Factors and levels

Factors	**Units**	**Levels**				
		$-\alpha$	-1	**0**	$+1$	$+\alpha$
Velocity (V)	pixel/(1/24sec)	0	8	16	24	44
	m/s	0	0.101	0.202	0.303	0.556
Distance from the screen (DFS)	m	-0.646	-0.321	0	$+0.317$	$+0.630$
Defocus blur amount (DBA)	pixel	0	1.6	2.4	4	6
	10^{-2}m	0	0.084	0.126	0.210	0.316

Thirty subjects who had no problem with watching 3D videos participated in the experiment. Their average age was23.0. Their eyesight was above 0.6, and the difference between their left and right eyesights was below 0.3. The experiment was performed as follows. We showed the subjects a video of two fish swimming, and asked them to select the fish on which they visually concentrated. The two fish on the video were made at the corresponding levels of the two different data points in Figure 2.1. A total of 105 combinations of two different points were produced from among 15 selected points, and the subjects evaluated the 105 videos. For each data point, the VAS was determined from the number of times the subject selected the fish that corresponded to that data point.

2.3.2 Analysis of Variance (ANOVA)

We performed Analysis of Variance (ANOVA) to determine if the velocity (V), distance from the screen (DFS), and defocus blur amount (DBA) were significant factors with respect to visual attention. We performed the analysis with the SPSS statistical program. We analyzed the combinations of the factors at level +1 and -1 for each factor. The data points according to these combinations are the black points in Figure

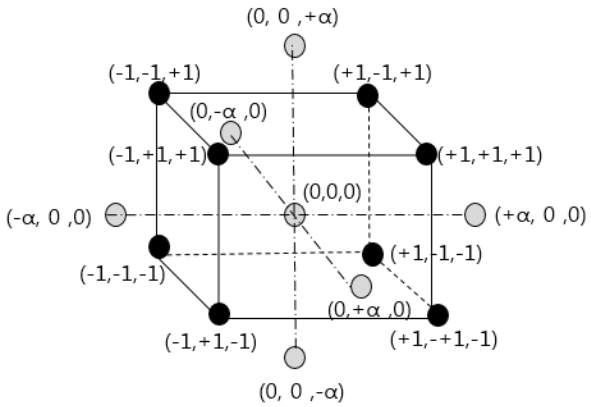

Figure 2.1: Data points.

Table 2.3: ANOVA table

Source	Sum of squares	Degree of freedom	Mean square	F	p-value
V	82.838	1	82.838	12.965	0.000
DFS	24.704	1	24.704	3.866	0.050
DBA	972.037	1	972.037	152.130	0.000
V×DFS	0.104	1	0.104	0.016	0.899
V×DBA	24.704	1	24.704	3.866	0.050
DFS×DBA	7.704	1	7.704	1.206	0.273
V×DFS×DBA	1.838	1	1.838	0.288	0.592
error	1482.367	232	6.390		
total	14539.000	240			

2.1. The analysis results are shown in Table 2.3. The p-value of the V and the DBA was less than 0.001, and the p-value of the DFS was 0.05. Therefore, all three factors were significant at a 5% significance level. That means all three factors are important factors of visual attention. Moreover, there was a significant interaction between the V and the DBA.

2.3.3 Response Surface Modeling (RSM)

In this chapter, we obtain the visual attention scores (VASs) at discrete data points and estimate the VAS model with respect to a continuous range. We define the scores as real numbers. For the discrete points, the experiment result values are between 0 and 15, but the scores at the other level combinations can be any value below 0 or above 15. We obtained the experiment result scores of the data points required for the RSM and performed regression with those scores. For the RSM, we used the Central Composite Design, which is widely used. All three factors could be included in the VAS model because they have been proven to be significant factors. In the case of the three factors, 15 data points were required for the Central Composite Design, and they are plotted in Figure 2.1. The average result score of each data point is shown in Table 2.4.

Following characteristics were obtained by average VAS distribution. Firstly, the visual attention increases when the object moves faster. And the faster the object is, the smaller the increment of the visual attention according to velocity increment is. Generally, when a fixed object and a slow-moving object exist on the same screen, viewers are highly likely to visually concentrate on the latter. However, when a fast-moving object and a object which moves faster exist together, their visual attention gap would be smaller than the first case. Fractional function form was selected for the V.

Secondly, the visual attention increases linearly when the object sticks out farther of the screen. However, the effect of this factor was less significant than the effects of

Table 2.4: Mean of visual attention scores at each data point

Data points	Levels of each factor			Mean VASs
	V (m/s)	DFS (m)	DBA (10^{-2}m)	
(-1,-1,-1)	0.101	-0.321	0.084	8.23
(-1,-1,$+1$)	0.101	-0.321	0.210	4.10
(-1,$+1$,-1)	0.101	$+0.317$	0.084	9.37
(-1,$+1$,$+1$)	0.101	$+0.317$	0.210	4.17
($+1$,-1,-1)	0.303	-0.321	0.084	8.90
($+1$,-1,$+1$)	0.303	-0.321	0.210	5.70
($+1$,$+1$,-1)	0.303	$+0.317$	0.084	9.77
($+1$,$+1$,$+1$)	0.303	$+0.317$	0.210	6.20
(0 , 0 , 0)	0.202	0	0.126	7.47
($-\alpha$, 0 , 0)	0	0	0.126	5.67
($+\alpha$, 0 , 0)	0.556	0	0.126	8.33
(0 ,$-\alpha$, 0)	0.202	-0.646	0.126	7.37
(0 ,$+\alpha$, 0)	0.202	$+0.630$	0.126	7.63
(0 , 0 ,$-\alpha$)	0.202	0	0	9.97
(0 , 0 ,$+\alpha$)	0.202	0	0.316	2.13

other two factors. Viewers generally encounter stuck out objects which are fast such as objects flying toward viewers. The reason that the viewers visually concentrate on theose kind of objects would be an effect of the V as well as such of the DFS. Affine function form was selected for the DFS.

Thirdly, the visual attention increases when the object gets in focus or less blurred. Gaussian filter standard deviations are used for the indicator of the blur amount, and the smaller the standard deviation is, the smaller the increment of the visual attention according to the standard deviation increment is. We selected quadratic function form for the DBA.

Finally, the visual attention significantly varies according to the DBA at a low V level, however, the VAS variance is relatively small at a high V level. Similarly, the VAS varies a lot according to the V at a high DBA level and varies less at a low DBA level. If one of the factor could attract the visual attention sufficiently, the effect of the other factor would be imperceptible. The multiplication term $DBA^2 \cdot \frac{1}{V+k}$ was added in response to the interaction of the V and the DBA.

We selected the function forms by considering the average VAS arrangement, graph form, and sum of the residual squares. The VAS model is as follows.

$$VAS = a \cdot DBA^2 + b_1 \cdot DBA + b_2 \cdot DFS + c \cdot \frac{1}{V+k} + d \cdot DBA^2 \cdot \frac{1}{V+k} + e \quad (2.6)$$

The coefficient and constant values drawn from the regression are as follows.

$$\begin{aligned} &a = 202901.342, \quad b_1 = -1677.090, \quad b_2 = 0.609, \quad c = -0.003, \\ &d = -146448.366, \quad e = 10.412, \quad k = 0.080 \end{aligned} \quad (2.7)$$

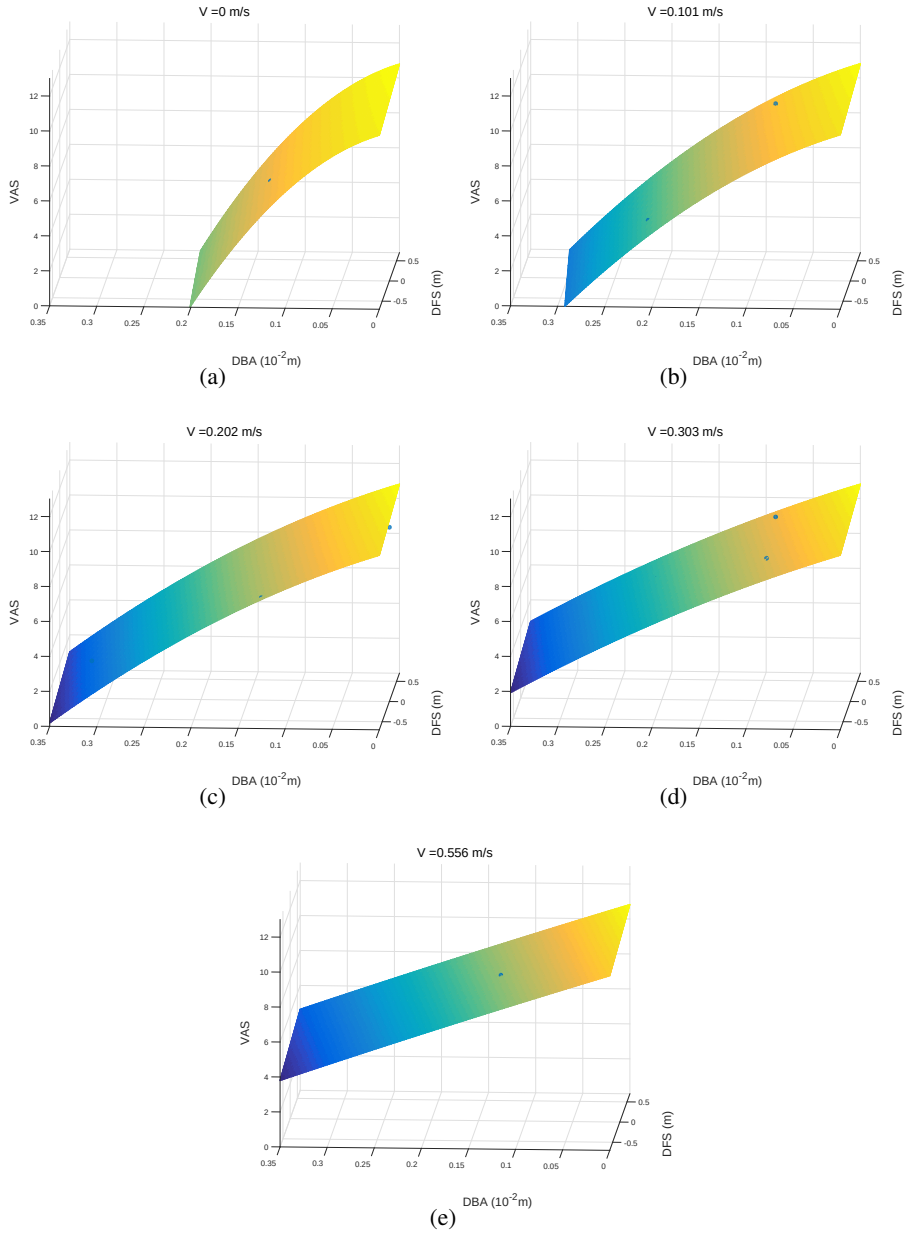

Figure 2.2: Visual Attention Score (VAS) graphs. (a), (b), (c), (d), (e) Velocity (V) is fixed to 0m/s, 0.101m/s, 0.202m/s, 0.303m/s, and 0.556m/s, respectively. Points are data points.

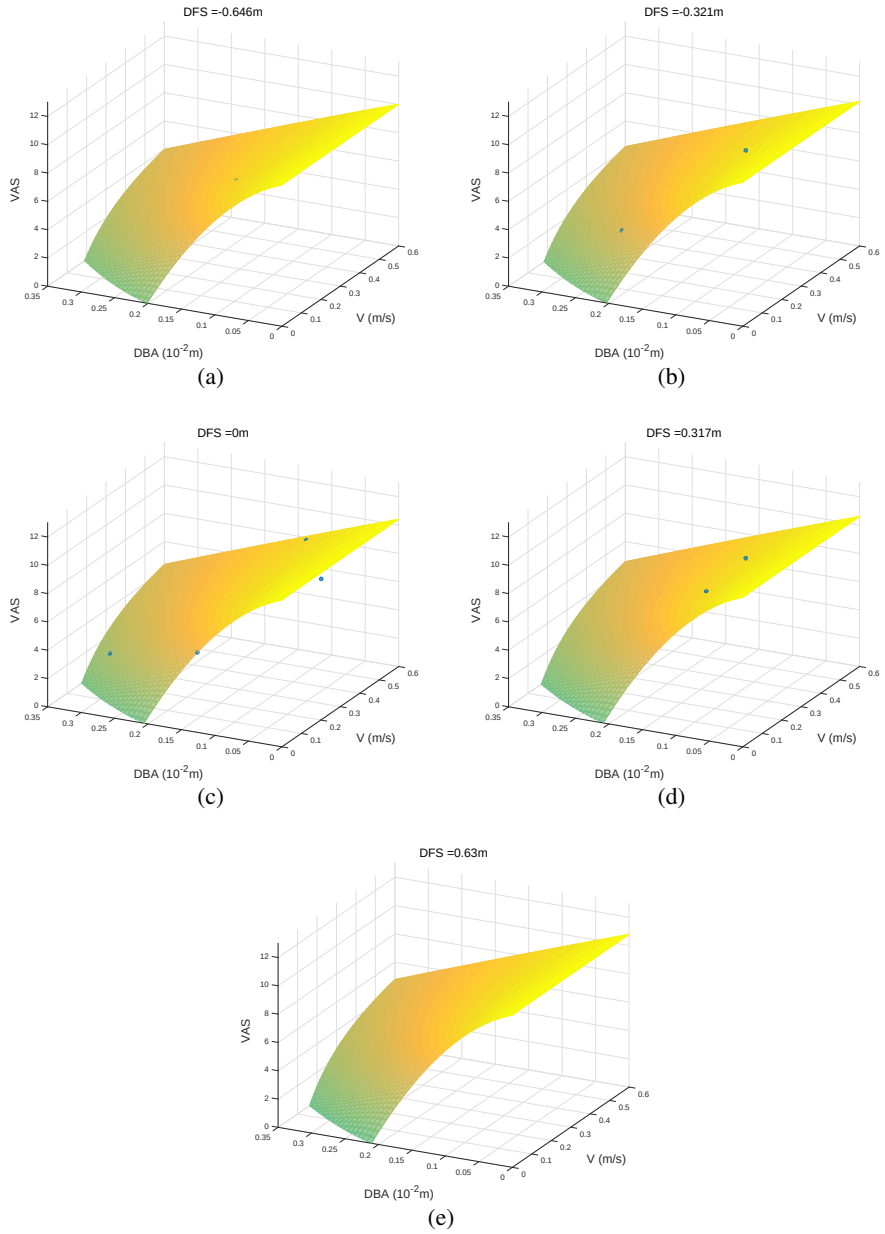

Figure 2.3: Visual Attention Score (VAS) graphs. (a), (b), (c), (d), (e) Distance From the Screen (DFS) is fixed to -0.646m, -0.312m, 0 m, $+0.317$m, and $+0.630$m, respectively. Points are data points.

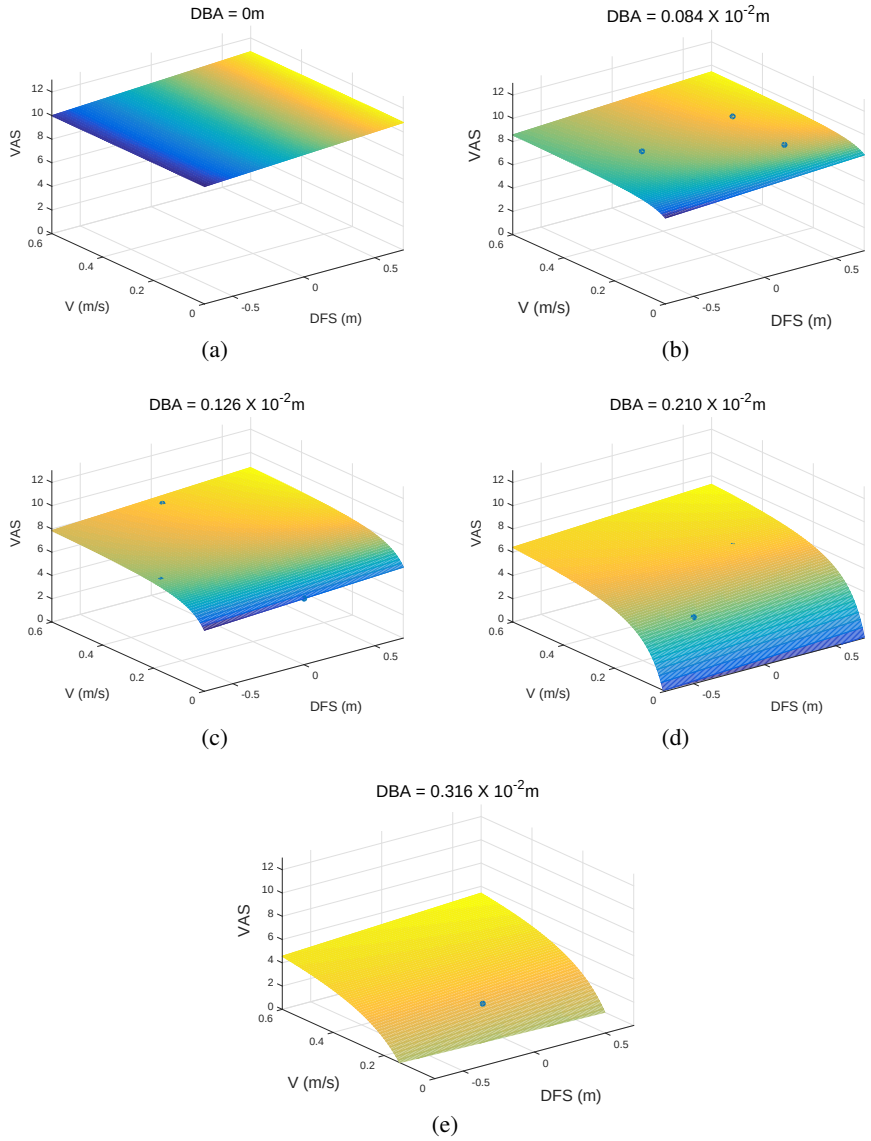

Figure 2.4: Visual Attention Score (VAS) graphs. (a), (b), (c), (d), (e) Defocus Blur Amount (DBA) is fixed to 0m, 0.084×10^{-2}m, 0.126×10^{-2}m, 0.210×10^{-2}m, and 0.316×10^{-2}m, respectively. Points are data points.

Figure 2.2 through 2.4 shows three dimensional graphs of the VAS model. Four axes are needed to represent the VAS according to the three factors. One factor should be fixed at a certain value to represent the model as a three dimensional graph. For instance, in Figure 2.2 (a), the V is fixed at 0 m/s, and the vertical and horizontal axes are for the DFS and the DBA, respectively. The up-down axis represents the VAS.

2.4 Visual Attention Probability and Its Application

2.4.1 Visual Attention Probability

We defined the Visual Attention Probability (VAP) of a pixel as the probability that a viewer will pay attention to the pixel from among all the image pixels. In this chapter, we calculate the VAP of each pixel in video frames using the earlier-obtained VASs. In a video frame, people will most likely pay attention to the pixel with the highest VAS. First, we mapped the VASs of all the pixels on the image. The mapping differed depending on the VAS range of the image. The mapping range was determined from the difference between the maximum VAS and the minimum VAS of the image. We mapped the VAS to make the mid-point of the maximum and minimum VASs 0.5, and the range, the mapping range. We increased the mapping range linearly according to the difference between the maximum VAS and the minimum VAS, when the difference was below the threshold, 2. When the difference was over the threshold, we saturated the mapping range to 1.

$$mapping\ range = \begin{cases} 1 & , if\ max(VAS) - min(VAS) > th \\ \frac{max(VAS) - min(VAS)}{th} & , otherwise \end{cases} \tag{2.8}$$

$$m(VAS(i,j)) = \frac{VAS(i,j)}{max(VAS) - min(VAS)} \cdot mapping\ range + \left[0.5 - \frac{max(VAS) + min(VAS)}{2}\right] \tag{2.9}$$

i and j are the vertical and horizontal pixel indexes, respectively, and m is the mapping function. We calculated the VAP using the mapped VAS, as follows.

$$VAP(i,j) = \frac{m(VAS(i,j))}{\sum_{i=1}^{height} \sum_{j=1}^{width} m(VAS(i,j))} \tag{2.10}$$

We performed the mapping with the standard mid-value, 0.5, so that the probabilities would not differ according to the bias. For example, assume that there are two images that consist of only two objects, and that one image has two objects with VASs 1 and 2, and the other image has two objects with VASs 9 and 10. If we calculate the probability without mapping, the difference between the probabilities of the objects will be much smaller in the latter image than the former image, even though the differences between the VASs are the same, i.e., 1. Therefore, we solved this problem by bringing the VAS values to around the standard mid-value.

We varied the mapping ranges depending on the VAS range of the image because of the human recognition ability. For example, when the difference between the maximum VAS and the minimum VAS of the images is so small that even the viewers cannot detect it, the VAP of the entire image should have little variance, but if we mapped the VAS from 0 to 1 for that image, the difference between the calculated probabilities would be significant. By observation, the probability that a viewer would concentrate on the pixel that had a maximum VAS value was very high when the difference between the maximum VAS and the minimum VAS was over 2. Therefore, we

saturated the mapping range over 2.

2.4.2 Application to Movie Frames

We applied the VAP to 3D movie frames. A 3D movie frame consists of a left image and a right image. To find the VAP, the V, DFS, and DBA of each pixel should be known. We used the following methods to obtain them.

In the case of V, we used Chan's method to calculate the motion from the left images of the target frame and the previous frame [9]. Chan increased the accuracy of the motion estimation by combining the block matching algorithm and the optical flow algorithm. In the experiment, we selected 30 pixels as the block size, and 60 pixels as the search range for 960 pixels with images.

To obtain the DFS, we need a disparity map. We used the software StereoTracer to obtain the disparity map from the left and right image of the frame. We calculated DFS from the disparity with the following equation [10].

$$DFS = \frac{d \cdot viewing\,distance}{-d + eye2eye} \tag{2.11}$$

DFS has a negative value when the pixel is behind the screen, and a positive value when the pixel is in front of the screen. d is the disparity distance on the screen converted from the number of disparity pixels. The unit of d is m. The viewing distance is the distance between the screen and the viewer, and for the experiment, we set it at 3 m. eye2eye is the distance between the left and right eyes, and for the experiment, we set it at 0.065 m, the average eye2eye of adults.

The defocus blur amounts for the edge pixels have been precisely determined in previous studies. But the blur amounts must be known for every pixels on an image. Some studies propagated the blur amount information on the edge pixels to an entire area, but the methods were not precise enough to be used in this paper. In this paper, we

devised a more accurate method of estimating the DBA using the distance information already known from the DFS.

We first obtained the in-focus distance, which is the distance from the viewer to the point most focused on, and assigned different DBAs depending on the distance from the in-focus distance. We obtained the in-focus distance from the fact that the defocus blur amount is lowest at the focused point. At first, we obtained the defocus blur amount at the edge pixels using Zhuo's method [11]. His method determines the standard deviation of the Gaussian kernel while assuming that the defocus blur is a Gaussian blur. Next, we divided the distance into sections. We used a 0.2m interval for each section, in the following experiment environment: viewing distance, 3 m, and screen width, 1.01 m. We calculated the mean blur amount in the edge pixels (MBAEP) in each section.

$$MBAEP(k) = \frac{\sum_{i=1}^{N_k} BAEP(i)}{N_k} \tag{2.12}$$

i is the index of the edge pixel in the k^{th} section; BAEP(i), the blur amount in the i^{th} edge pixel; and N_k, the number of edge pixels in the k^{th} section. When the standard deviation of the MBAEPs was smaller than the threshold, we considered the blur amount difference very small in an entire image and assigned the average value of the MBAEPs to all the pixels in the image. We used a 0.15 pixel as the threshold for a 960-pixel width image. When the standard deviation was larger than the threshold, we found the section with the minimum MBAEP. The focused point would be around that section. To find a more accurate in-focus distance, we scanned around that section with a 0.2m window and found the window that had the minimum MBAEP. We determined the in-focus distance as the center distance of that window.

We decided the function of DBA according to the distance from the in-focus distance as the modified-Gaussian form. It is shown in Equation 2.13.

$$DBA(i,j) = \begin{cases} mean(MBAEP(k)) & , \ if\ std(MBAEP(k)) < threshold \\ max(MBAEP(k)) - \Big[(max(MBAEP(k)) - min(MBAEP(k))) & \\ \cdot \exp \frac{-(DFS(i,j) - infocus\ distance)^2}{2 \times \sigma_f(i,j)} \Big] & , \ otherwise \end{cases} \quad (2.13)$$

$$\sigma_f = 0.6 \cdot DOF(viewing\ distance - DFS(i,j)) \quad (2.14)$$

The DBA function is like an upside-down Gaussian function, and its maximum value is max(MBAEP(k)), and its minimum value, min(MVAEP(k)). The term that corresponds to the standard deviation differs depending on the distance. It was determined as proportional to the Depth of Field (DOF), which was calculated assuming that each distance is the in-focus distance. The proportional coefficient was determined to have been 0.6, considering the shape of the graph and the DOF. The DOF is the distance between the nearest and farthest objects in a scene that appears acceptably sharp in an image, and it changes depending on the in-focus distance. When the focused point comes nearer to the viewer, the DOF becomes smaller. The DOF is obtained using the following formula, according to Greenleaf [12].

$$H = \frac{f^2}{N_c} + f \quad (2.15)$$

$$D_n(s) = \frac{s(H-f)}{H+s-2f}, \qquad D_f(s) = \frac{s(H-f)}{H-s} \quad (2.16)$$

$$DOF(s) = |D_n(s) - D_f(s)| \quad (2.17)$$

H is the hyperfocal distance; f, the lens focal distance; s, the focus distance; D_n, the near distance for acceptable sharpness; D_f, the far distance for acceptable sharpness;

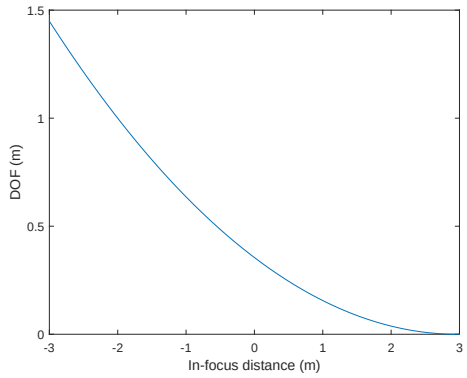

Figure 2.5: In-focus distance and DOF.

N, the f-number; and c, the circle of confusion. In the experiment, the DOF was calculated based on the canon 7D (circle of confusion, 0.019 mm; focal length, 55 mm; and f-number, 3.2). Figure 2.5 represents the DOF values according to the in-focus distance. Figure 2.6 presents the DBA values according to the distance, in case the in-focus distance is 1 and -1, respectively.

The VAP is calculated from the V, DFS, and DBA, which are obtained as previously described. Figure 2.7 through 2.16 present the results. (a) shows the left image; (b), VAP; (c), V; (d), DFS; and (e), DBA. For each factor, a number is multiplied by the values shown well by the image. In the images, the color of the pixel with a large VAP, V, DFS, and DBA appears close to white. In the case of test sequence 3, presented in Figure 2.9, there was a slight difference in the background and the object. Therefore, all the pixels in the image had a constant DBA.

Cases that the effects of multiple factors conflict are mainly selected for the examples. In these cases, visual attention would be changed according to the levels of the factors and it is hard to derive visual attention model which is robust on such cases without an experiment performed with the factor level combinations. Our model is estimated by above-mentioned experiment and we obtained reasonable results on the

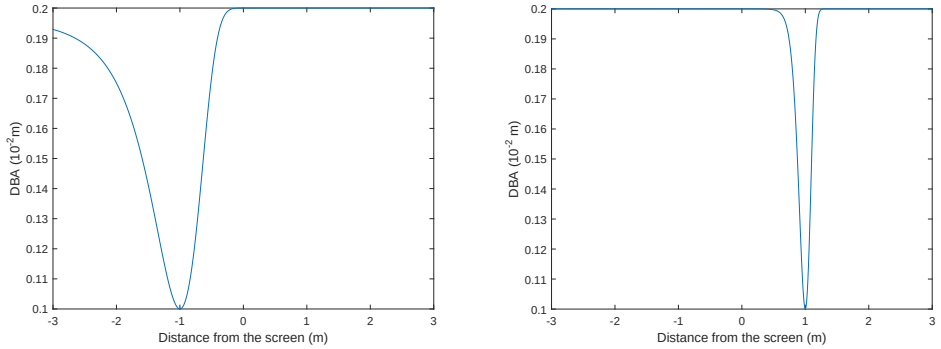

Figure 2.6: Example graphs of DBA according to distance from the screen in the specific in-focus distances. (a) when in-focus distance is 1, (b) when in-focus distance is -1. When max(BAEP(k)) = 0.2×10^{-2}m, min(BAEP(k)) = 0.1×10^{-2}m.

tests.

In the case of Fig. 2.11, from the image (a), we expect that the viewers would be concentrated on the person on the left which was the salient object when shooting, even though person on the right stuck out. And the VAP map corresponds with the expectation. In Fig. 2.8, the V and the DBA which affect visual attention a lot were distributed to the different parts; the hand and the body, so the VAPs of the parts are expected to come out similar. In Fig. 2.7, head of the person in the center moved, stuck out and got in-focus, so the VAP of the head are expected to be the biggest one. In Fig. 2.9, two people in the center moved fast, therefore the VAPs are expected to be the biggest ones. Two people in the edge moved slow but stuck out, therefore the VAPs would be quite big.

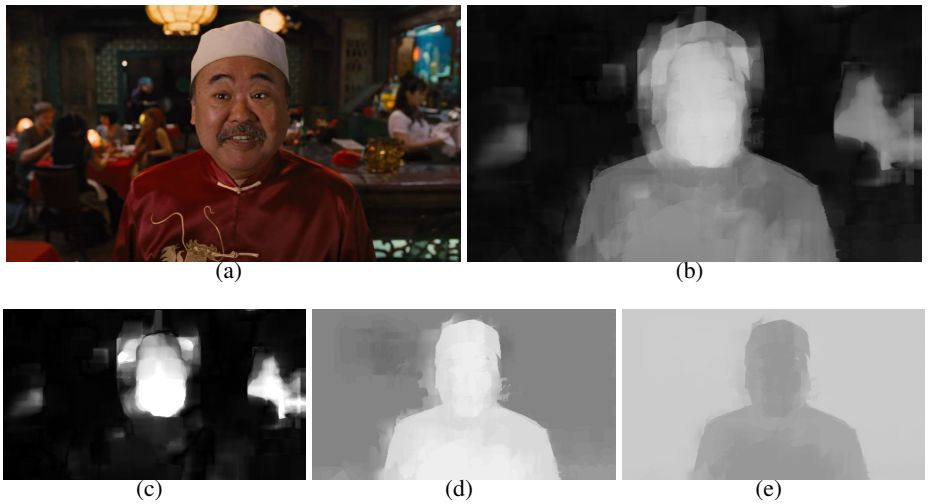

Figure 2.7: Result images for test sequence 1. (a) Test frame's left image. (b) VAP map. (c) V map. (d) DFS map. (e) DBA map. White represents a big value.

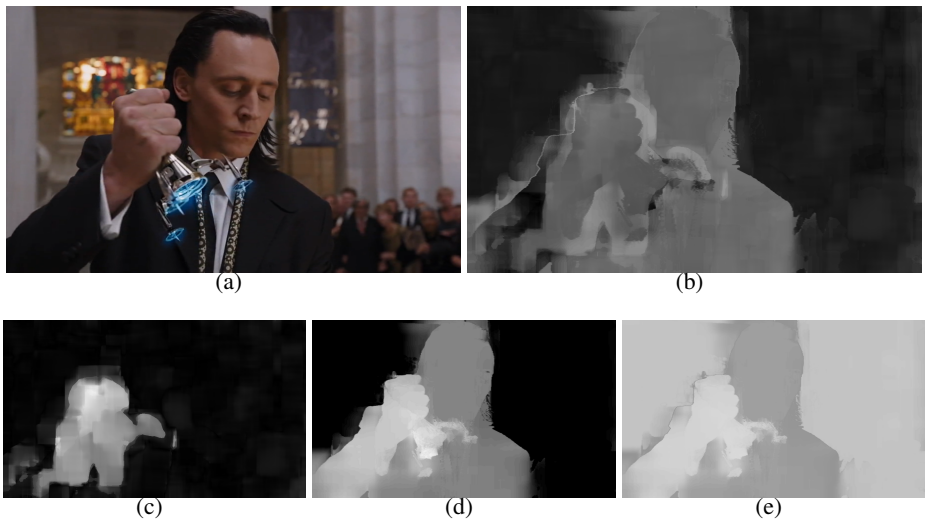

Figure 2.8: Result images for test sequence 2. (a) Test frame's left image. (b) VAP map. (c) V map. (d) DFS map. (e) DBA map. White represents a big value.

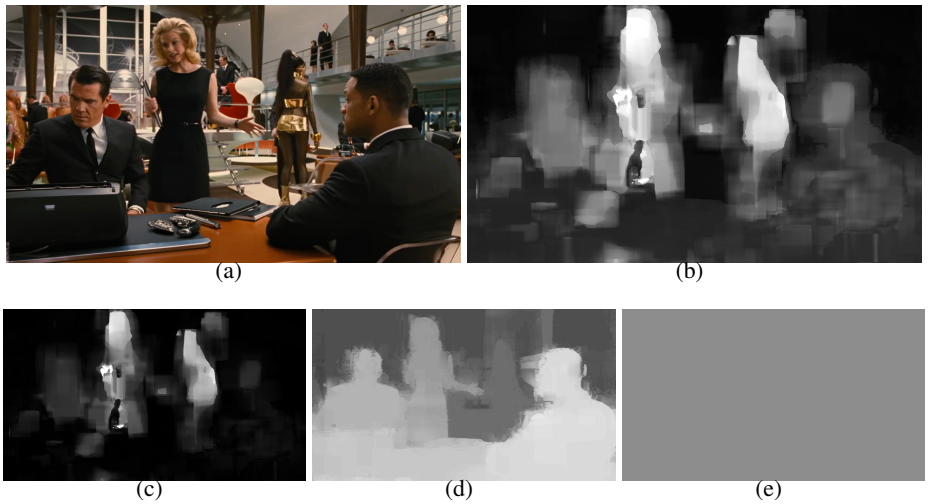

Figure 2.9: Result images for test sequence 3. (a) Test frame's left image. (b) VAP map. (c) V map. (d) DFS map. (e) DBA map. White represents a big value.

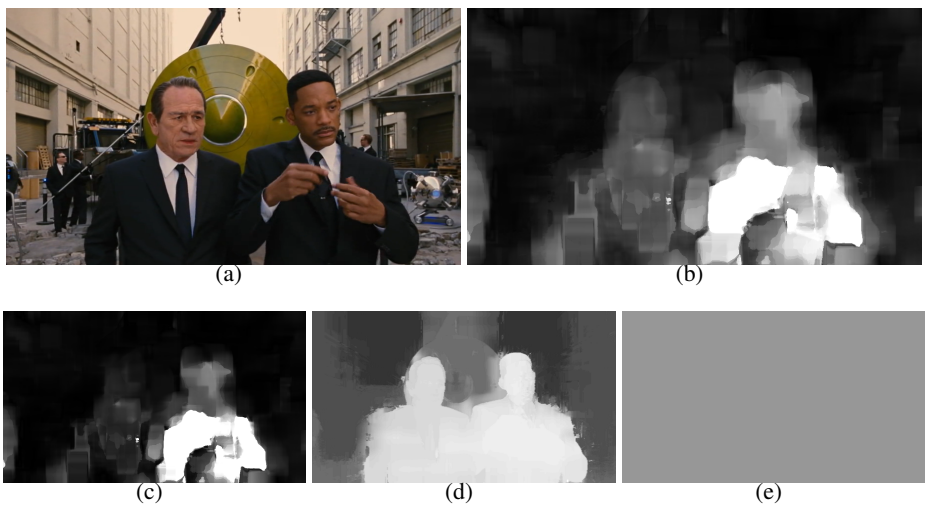

Figure 2.10: Result images for test sequence 4. (a) Test frame's left image. (b) VAP map. (c) V map. (d) DFS map. (e) DBA map. White represents a big value.

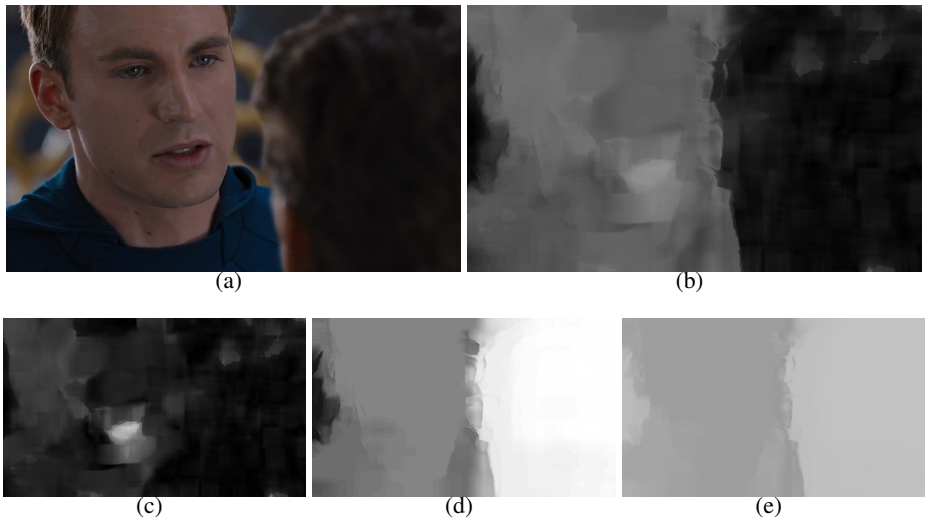

Figure 2.11: Result images for test sequence 5. (a) Test frame's left image. (b) VAP map. (c) V map. (d) DFS map. (e) DBA map. White represents a big value.

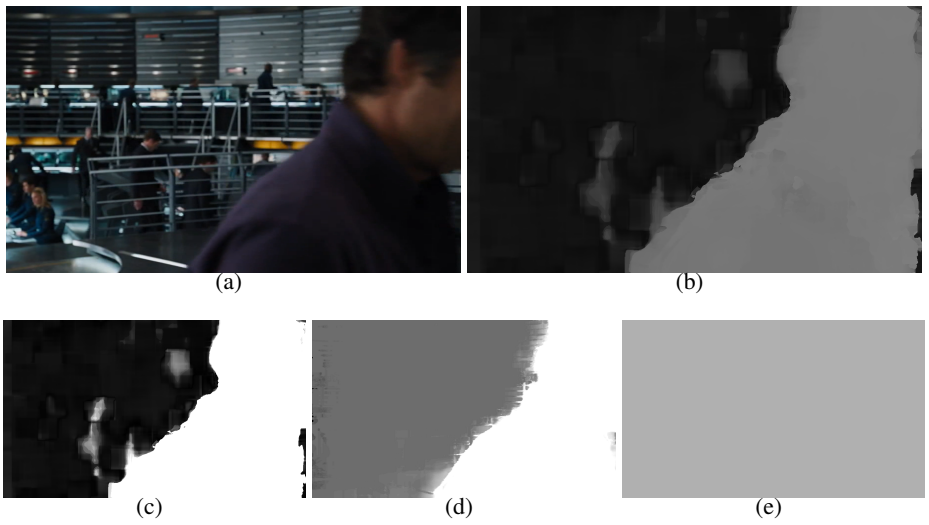

Figure 2.12: Result images for test sequence 6. (a) Test frame's left image. (b) VAP map. (c) V map. (d) DFS map. (e) DBA map. White represents a big value.

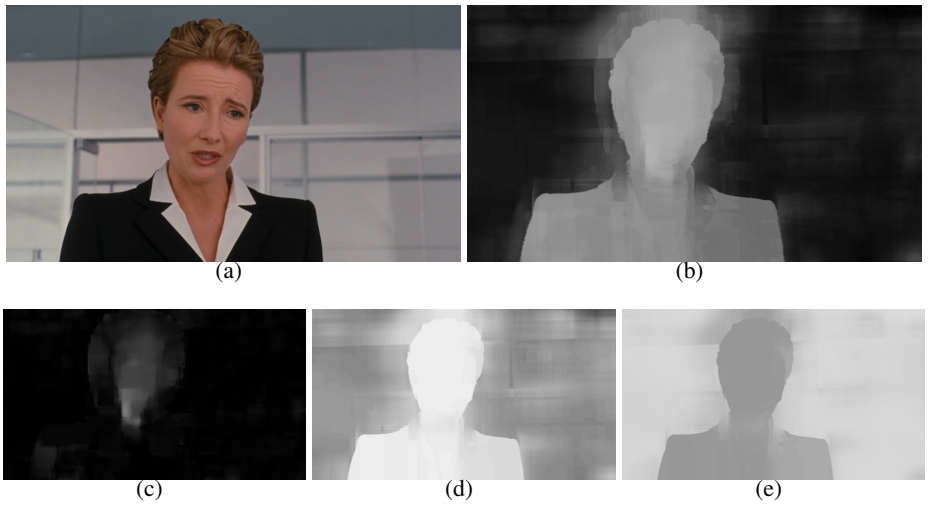

Figure 2.13: Result images for test sequence 7. (a) Test frame's left image. (b) VAP map. (c) V map. (d) DFS map. (e) DBA map. White represents a big value.

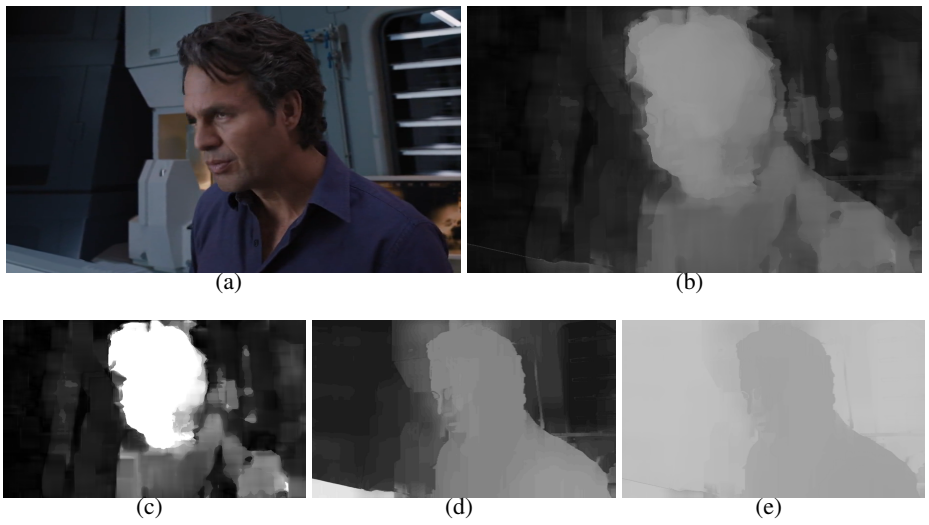

Figure 2.14: Result images for test sequence 8. (a) Test frame's left image. (b) VAP map. (c) V map. (d) DFS map. (e) DBA map. White represents a big value.

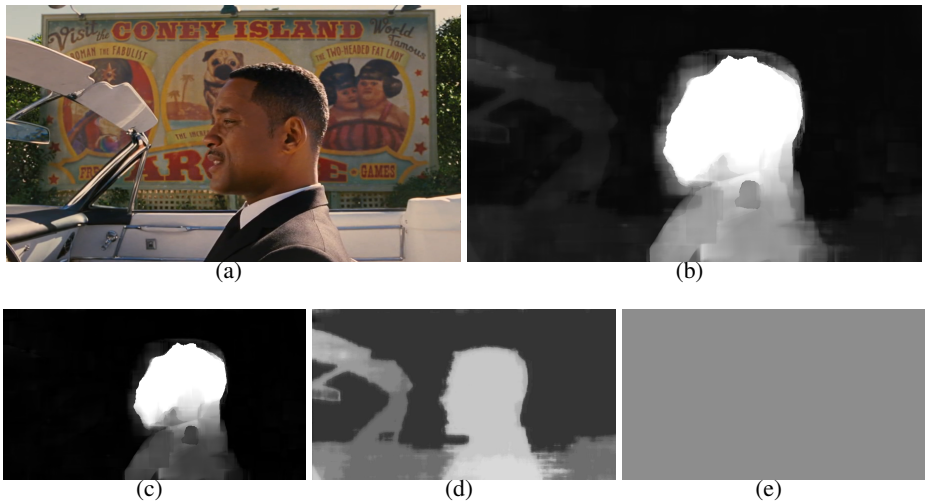

Figure 2.15: Result images for test sequence 9. (a) Test frame's left image. (b) VAP map. (c) V map. (d) DFS map. (e) DBA map. White represents a big value.

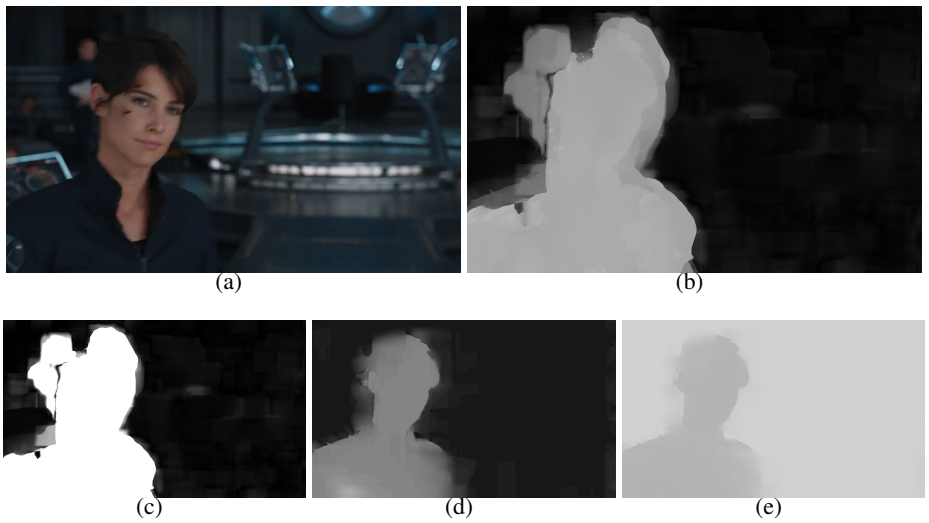

Figure 2.16: Result images for test sequence 10. (a) Test frame's left image. (b) VAP map. (c) V map. (d) DFS map. (e) DBA map. White represents a big value.

Chapter 3

3D Effect Perception Measurement Using the VAP

3.1 Previous Studies

In this paper, we estimate the 3D effect perceived by the viewers of different scenes of stereoscopic videos.Many studies have focused on visual fatigue as a criterion of 3D video evaluation, but there have been few studieson3D effect perception. Choi measured the quality of the 3D effect through a depth image histogram [14]. He argued that the bigger the depth variety is, the more details are represented and the more natural the object edges seem, in the same image. He estimated that the depth variety is big when the depth image histogram is similar to the uniform distribution. Kim measured the 3D effect from the Gradient Magnitude Average (GMA) of the depth image [15]. He argued that the gradient operation extracts the spatial variations of the depth image, and that the spatial variation of the depth image provides meaningful information on the front-back relationship of objects.

$$\bigtriangledown I = (I_x, I_y) = \Big(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}\Big) \tag{3.1}$$

$$\| \bigtriangledown I\| = \sqrt{I_x^2 + I_y^2} \tag{3.2}$$

$$GMA = \frac{\sum_{x=1}^{height} \sum_{y=1}^{width} \| \bigtriangledown I\|}{height \cdot width} \tag{3.3}$$

$\bigtriangledown I$ represents the gradient of a depth image, and $\| \bigtriangledown I\|$, the gradient magnitude. In Kim's experiments, he produced eight different depth images for the same scene, and measured the perceived 3D effects of the 3D images produced by the depth images. However, it is also possible to measure the perceived 3D effect for different scenes using the GMA.

In this paper, we propose a method that more accurately measures the 3D effect perception by applying VAP to the gradient method.

3.2 Gradient Method with the VAP

A viewer accepts information around a visually concentrated point. A 3D effect perceived by viewers would be significantly affected by the depth information around a visually concentrated point. In this paper, we gave weights to gradient magnitudes of the pixels around a visually concentrated point depending on the Visual Sensitivity Kernel (VSK). By applying the VAP to each pixel, we derived the following equation for an entire image. We call this measurement the "Weighted Gradient Magnitude by the VAP (WGMVAP)."

$$WGMVAP = \sum_{i,j=1}^{height,width} \sum_{x,y=1}^{height,width} VAP(i,j) \cdot VSK(x-i, y-j) \cdot \| \bigtriangledown D(x,y)\| \tag{3.4}$$

We used disparity images obtained from the left and right images of stereoscopic videos instead of Kim's depth image. D refers to the disparity image; $\| \bigtriangledown D(x,y) \|$, the gradient magnitude of the disparity image; i and x, the vertical indexes of an image; and j and y, the horizontal indexes of an image. Equation 3.4 can be expressed again as Equation 3.5 and 3.6 using convolution.

$$weight = VAP \otimes VSK \tag{3.5}$$

$$WGMVAP = \sum_{x,y=1}^{height,width} weight(x,y) \cdot \| \bigtriangledown D(x,y) \| \tag{3.6}$$

The VSK is determined as follows. In the medical field, Goldmann studied visual sensitivity to light [1]. He found that the visual sensitivity rises steeply within a two-degree angle from a visually fixed point. We could not directly use the figures for sensitivity from his study to this paper because his research was for light. However, we refer to the range in which the sensitivity increases steeply. In the case of a 3D space, a steeply increasing range would be narrower than a plane case. If a human focused on an object at a certain distance, the visual sensitivity would fall with respect to the objects at another distance. In the experiment, we estimated the angular range in which the sensitivity increases steeply as 1 degree, and made a kernel. We converted this angular range to the distance range on the screen using the following equation: $viewing\ distance \cdot tan(1^\circ)$. We selected a cone as the form of the kernel, and we decreased it linearly from a point visually fixed at on the vertical and horizontal axes, respectively. Figure 3.1 shows a graph of the kernel.

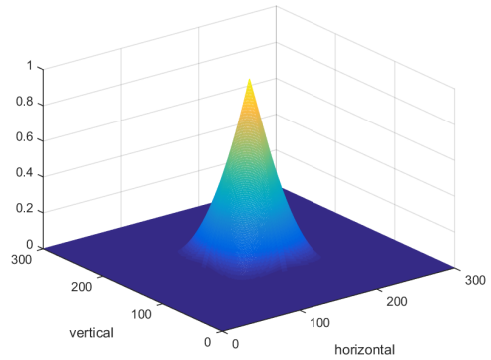

Figure 3.1: Visual Sensitivity Kernel (VSK).

3.3 Experiment Results

To check the accuracy of the 3D effect perception measurement, we compared the measurement results with the results of the subject evaluation. For the evaluation, we asked the subjects to evaluate their perceived 3D effect with respect to 20 video clips that were each one to two seconds long. The evaluation scores were 1 (very small), 2 (small), 3 (middle), 4 (big), and 5 (very big), and we showed two standard videos that were relevant to 1 and 5. We asked the subjects to evaluate the 3D effect in the last parts of the videos. Video clips were extracted from four well-known 3D movies: Men in Black 3, Gravity, Life of Pi, and Avengers. Thirty-five subjects who had no problem with watching 3D videos participated in the experiment. Their average age was 23.1. All of them had an eyesight above 0.6, and the difference between their left and right eyesights was below 0.3.

We compared the performance of the existing method, GMA [15] and the proposed method, WGMVAP. In addition, we applied Kim's visual attention model, 3DVA [2] to the 3D effect perception measurement and obtained the performance of the measurement. We call this measurement the "Weighted Gradient Magnitude by 3DVA (WGM3DVA)." $3DVA^n$ in the equations below denotes the normalized 3DVA model

[2].

$$WGM3DVA = \sum_{i,j=1}^{height,width} \sum_{x,y=1}^{height,width} 3DVA^n(i,j) \cdot VSK(x-i, y-j) \quad (3.7)$$

$$3DVA^n(i,j) = \frac{3DVA(i,j)}{\frac{1}{height \cdot width} \sum_{i,j=1}^{height,width} 3DVA(i,j)} \quad (3.8)$$

Figure 3.2 presents the experiment results. Figure 3.2 (a) shows the scatter plot whose horizontal axis is the average value of the subjective evaluation and whose vertical axis is the result derived from the GMA. Each point represents a video clip. Figure 3.2 (b) shows the scatter plot whose vertical axis is the result derived from the WGM3DVA. Figure 3.2 (c) shows the scatter plot whose vertical axis is the result derived from the proposed method, WGMVAP. We confirmed that the performance improved after the application of the visual attention model in terms of the coefficient of determination R^2. In addition, we confirmed that the performance was higher when the VAP was applied than when the 3DVA was applied. R^2 for the GMA, WGM3DVA, WGMVAP were 0.456, 0.591, and 0.612, respectively.

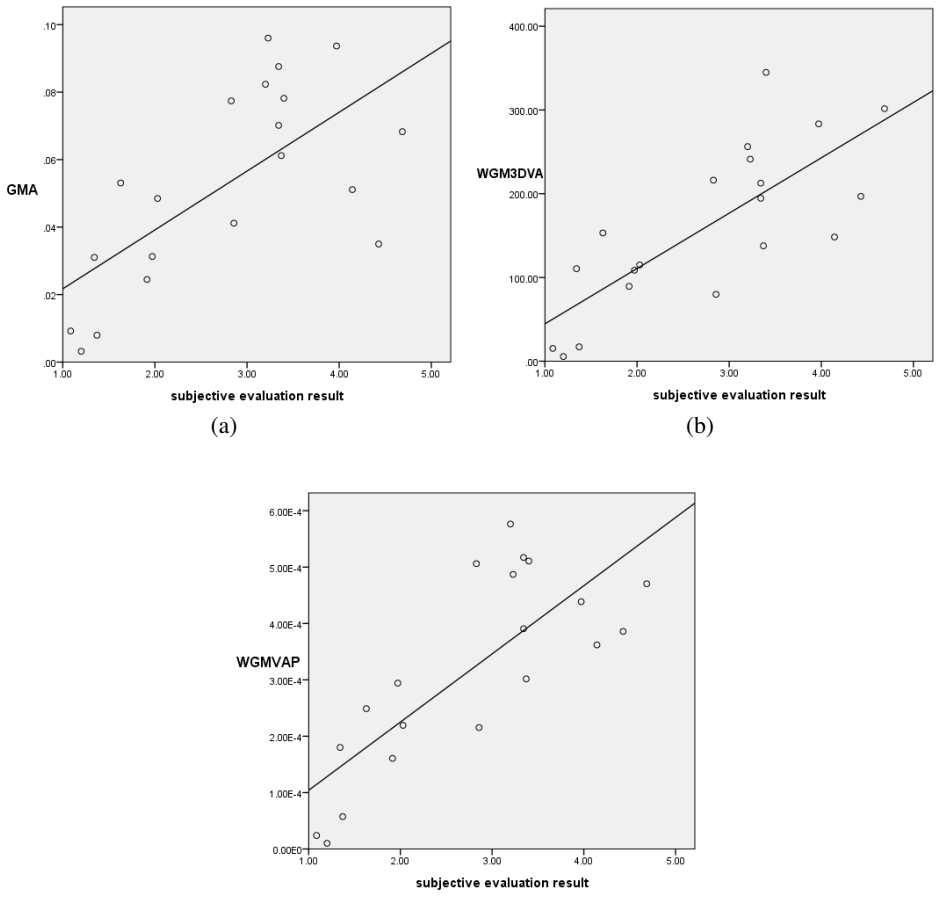

Figure 3.2: (a) Scatter plot of GMA and subjective evaluation result. Linear optimum function : $y = 1.74 \times 10^{-2}x + 4.28 \times 10^{-3}$, $R^2 = 0.456$. (b) Scatter plot of WGM3DVA and subjective evaluation result. Linear optimum function : $y = 6.61 \times 10x - 2.15 \times 10$, $R^2 = 0.591$. (c) Scatter plot of WGMVAP and subjective evaluation result. Linear optimum function : $y = 1.21 \times 10^{-4}x - 1.69 \times 10^{-5}$, $R^2 = 0.612$.

Chapter 4

Conclusion

In this paper, we estimated the visual attention probability (VAP) model for videos using the statistical experiment design. From the ANOVA, we confirmed that the velocity (V), distance from the screen (DFS), and defocus blur amount (DBA) are the factors that significantly affect people's visual attention. In addition, from the RSM, we estimated the visual attention score (VAS) form in continuous factor ranges. From this VAS model, we calculated the VAPs of the image pixels.

Next, we applied the VAP model to the 3D effect perception measurement for stereoscopic videos. We proposed the application method to improve the performance of Kim's measurement using the gradient method. From the experiment results, we confirmed that the proposed method improved the performance of the existing measurement.

In this paper, we used 15 data points for the response surface modeling (RSM), but a more accurate model could be made with more data points. More studies are needed to determine the shape of the visual sensitivity kernel (VSK). In addition, if more accurate measurement methods for the velocity and disparity are used, especially around the edge pixels, a more accurate VAP could be obtained.

Bibliography

[1] J. F. DUANE, *Duane's Ophthalmology* , Lippincott Williams & Wilkins, 2006.

[2] 김동현, 손광훈, "3 차원 시각 주의 모델과 이를 이용한 무참조 스테레오스코픽 비디오 화질 측정 방법," *전자공학회논문지*, 제51권, 제4호, 786-798쪽, 2014년 4월.

[3] B. Mendiburu, *3D movie making : Stereoscopic digital cinema from script to screen*, Elsevier, America, 2009.

[4] L. Itti, C. Koch, and E. Neibur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, vol. 20, no.11, pp. 1254-1259, November 1998.

[5] Y. Park, B. Lee, W. s. Cheong, and N. Hur, "Stereoscopic 3D visual attention model considering comfortable viewing," in *Proceedings of IET Conference on Image Processing* , London, United Kingdom, July 2012. pp. 1-5.

[6] D. Chai and A. Bouzerdoum, "A Bayesian approach to skin color classification in YCbCr color space," in *Proceedings of TENCON* , Kuala Lumpur, Malaysia, September 2000. vol. 2, pp. 421-424

[7] 임채영, 박세근, "통계적 실험계획법을 이용한 SOG 평탄화 공정의 최적화," *한국진공학회지*, 제1권, 제1호, 198-205쪽, 1992년 2월.

[8] 박성현, *현대실험계획법* , 민영사, 서울, 2012.

[9] S. H. Chan and D. T. V$\tilde{o}$, and T. Q. Nguyen, "Subpixel motion estimation without interpolation," in *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing* , Dallas, TX, America, March 2010. pp. 722-725

[10] L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3D TV," *IEEE Transactions on Broadcasting.*, vol. 51, issue. 2, pp. 191-199, June 2005.

[11] S. Zhuo and T. Sim, "Defocus map estimation from a single image," *Pattern Recognition.*, vol. 44, no.9, pp. 1852-1858, March 2011.

[12] Greenleaf and R. Allen, *Photographic Optics*, The MacMillan Company , New York, 1950.

[13] 최지훈, 이강규, 김아란, 김종옥, "깊이 영상의 히스토그램을 이용한 스테레오 영상의 입체감 측정," *대한전자공학회 2011년 추계종합학술대회*, 2011년 11월. pp. 505-506.

[14] J. H. Choi, J. W. Kim, and J. O. Kim, "Stereoscopic depth perception measurement for 2D/3D converted contents," in *Proceedings of IEEE Conference on Consumer Electronics*, Tokyo, Japan, October 2012. pp. 498-501.

[15] 김재우, 최지훈, 김종옥, "깊이 영상의 gradient 기반 스테레오 영상 입체감 측정 기법," *대한전자공학회 2011년 추계종합학술대회*, 2011년 11월. pp. 511-512.

[16] J. W. Kim, J. H. Choi, and J. O. Kim, "Stereoscopic depth perception measurement using depth image gradient," in *Proceedings of International Conference on Awareness Science and Technology*, Seoul, Korea, August 2012. pp. 141-145.

[17] H. Sohn, Y. J. Jung, S. I. Lee, and H. W. Park, "Attention model-based visual comfort assessment for stereoscopic depth perception," in *Proceedings of Inter-*

national Conference on Digital Signal Processing, Corfu, Greece, July 2011. pp. 1-6.

[18] S. Lee, J. Kim, and S. Choi, "Real-time tracking of visually attended objects in virtual environments and its application to LOD," *IEEE Transactions on Visualization and Computer Graphics.*, vol. 15, issue.1, pp. 6-19, January 2009.

[19] 박진희, "입체영상에서 깊이의 정도가 시각적 주의에 미치는 영향," *디지털디자인학연구*, 제10권 제2호, 441-450쪽, 2010년 4월.

[20] J. Choi, D. Kim, B. Ham, S. Choi, "Visual fatigue evaluation and enhancement fro 2D-plus-depth video," in *Proceedings of IEEE Conference on Image Processing*, Hong Kong, September 2010. pp. 2981-2984.

국문 초록

시청자들은 화면상 시각이 집중된 곳 주변의 정보를 영향력 있게 받아들일 가능성이 크다. 이러한 사실을 이용하여 최근 연구들은 시각 주의 모델을 영상 제작 및 평가 방법에 이용하고 있다. 본 연구에서는 실제로 사람들의 시각 주의도가 어떠한 인자에 영향을 많이 받는지, 또 시각 주의 확률 모델은 구체적으로 어떠한 형태가 되는지 통계적 실험 계획법을 이용하여 추정하였다. 분산 분석법을 이용하여 속도, 화면으로부터의 거리, 비초점 흐림 정도가 시각 주의에 영향을 미치는 유의한 인자인 것을 확인하였고, 반응 표면 계획법을 이용하여 이 세 가지 인자들에 따른 시각 주의 점수 모델을 도출하였다. 이 시각 주의 점수 모델로부터 이미지 각 픽셀의 시각 주의 확률을 구하였다. 본 연구의 뒷부분에서는 시각 주의 확률 모델을 기존의 기울기 기반 3 차원 영상의 입체감 추정 방법에 적용하는 방법을 제안하였다. 화면 상에서 시선을 집중할 확률이 큰 부분에 높은 비중을 둠으로써 시청자가 느끼는 입체감을 더욱 정확하게 측정할 수 있도록 하였다. 제안된 방법의 성능을 검증하기 위해 주관적 평가를 실시하여 피실험자들이 느끼는 입체감과 제안된 방법으로부터 도출한 결과를 비교하였다. 실험 결과 제안된 방법이 기존의 방법에 비해 성능이 향상된 것을 확인하였다.

주요어: 시각 주의 확률, 3차원 동영상, 통계적 실험 계획법, 입체감 측정, 디포커스 흐림

학번: 2013-20759